Detection of Anomalies in Accounting Data Using Benford's Law: Evidence from India

Ramesh Chandra Das (Corresponding author)

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Tel: 91-900-288-3115   E-mail: rameshchandradas99@gmail.com

Dr. Chandra Sekhar Mishra

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Tel: 91-322-228-2318   E-mail: csmishra@vgsom.iitkgp.ernet.in

Dr. Prabina Rajib

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Tel: 91-322-228-3886   E-mail: prabina@vgsom.iitkgp.ernet.in

Abstract

This study uses the financial accounting data to examine if they depart from Benford's Law. Using large sample of Indian public listed companies, the study conducts an analysis of the "first digit analysis", "second digit analysis", and "first two digit analysis "of test variables such as total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, cost of goods sold, cash, EBIT, direct tax, indirect tax. The initial results find that most of the variables have significant deviation from Benford's Law distribution. Further analyses indicate that business

group firms indulge more data anomalies than standalone firms and small size firms have more data anomalies than large size firms in Indian context.

Keywords: Accounting data, Benford's law, Data anomalies

1. Introduction

Financial reporting covers all most every aspects of business operation. Financial statements are used by the stakeholders for assessing the present financial strength and viability in the future and to draw inferences about the business strategies. The annual report and the quarterly report that a company publishes are definitive instruments for the investor community to understand the performance and compliance with generally accepted accounting principles (GAAP). These are statutory requirements and contain a lot of relevant numbers. The analysts use the numbers for prediction; thegovernment uses those for tax calculation; investors use the numbers for performance measurement and bankers use the numbers for measuring the financial strength before granting loans. With such importance attached to the numbers disclosed in the financial reports, there can be chances that the certain financial numbers are not genuinely reported. There are instances of companies fudging financial numbers. Vishal Mega Mart-India (Note 1) (2007)a retailer indulged in cooking books by Rs.70 crore, Subhiksha (Note 2) (2008) inflated revenue figures and fudged accounting transactions by Rs.70 crore, Reebok (Note 3) (2011) falsified accounts, and apparently over-invoiced the sales by Rs.870 crore, DLF (Note 4) (2012) inflated sales by at least Rs.11,236 crore. Hence, there is a need to study the data anomalies if any in the reported financial results of the companies.

In recent years, Benford's Law has been found as a sophisticated statistical technique in fraud detection process. It is used widely as a hand tool by auditors in accounting and auditing fraud detection. This law holds a great promise for fraud detection process in the digital analysis (Coderr, 2000). Benford's Law is an expected frequency distribution of leading digits in many (but not all) real-life sets of numerical data. The law states that smaller digits occur more frequently than larger digits. For example in a set of numerical data, the number 1 would appear as the most significant digit about 30% of the time, while digit 9 would appear less than 5% of the time (Figure 1).
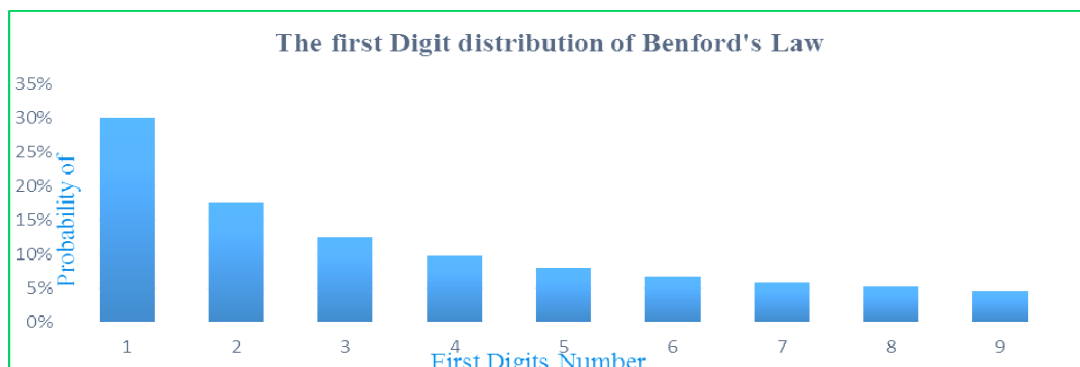

Figure 1. The first digit distribution of Benford's law

Researchers in social science as well as natural science have used the Benford's Law. Many

researchers have verified for themselves that Benford's Law is widely obeyed. In practice, the accuracy of Benford's Law has been experimented in many fields, such as accounting fraud detection (Nigrini, 1996), electricity bills, street addresses, stock prices, macroeconomic data like population numbers, death rates, lengths of rivers (Rauch et al. 2011), scientific fraud detection (Diekmann, 2007).

The law came out in 1881, when the American astronomer and mathematician (Newcomb, 1881) published the first known article in the American Journal of Mathematics, what is known as Benford's Law. He observed that library copies of books of logarithms were considerably more worn in the beginning pages which dealt with low digits and progressively less worn on the pages dealing with a higher digit. Newcomb (1881) provided no theoretical explanation which went virtually unnoticed. A year later 50 years, in 1938, Frank Benford came with article "The Law of Anomalous Numbers" published in American Philosophical society. That article is known as Benford's Law (hereafter Benford's Law). The law is named after Frank Benford's, who noticed that the first few pages of his table of common logarithms were more worn than the later pages (Benford, 1938). Benford's Law is an expected distribution of digits in tabulated data, such that when the data are ranked from smallest to largest they form a geometric sequence. Raimi (1976) observed that almost all sequences defined by linear recursions follow Benford's Law. He also observed that if a sequence does not conform to Benford's Law, there is noconstant that multiplication would cause it to become a Benford's set. As pointed out by Nigrini (1996), "conformity of a data set to Benford's Law does not necessarily imply naturalness, but nonconformity should raise some level of suspicion.

In this study, we introduce the first digit, second digit and first two digit tests of the financial statement items like total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, cost of goods sold, cash, earnings before interest and taxes (EBIT), direct tax and indirect tax in case of Indian companies for examining the reliability of data using Benford's Law distribution.

These accounting items are the most important reporting financial figures that are used for calculation of financial ratios. In particular, these are used for calculation of profitability ratio like return on assets (ROA), return on equity (ROE), and return on sales (ROS). The accounting items taken for this study are a mix of transactional level data and set of numbers that are from a mathematical combinations. Durtschi et al. (2004) explored the data that are likely to conform to Benford's Law and the conditions under which a Benford's Law is likely useful. The data should come from mathematical combinations, such as accounts receivable (number sold * price), accounts payable (number bought * price). The data should be transaction level data such as sales, and expenses. Alali and Romero (2013) find the double entry accounting, and cumulative natures of balance sheet items are manipulated.

The rest of the paper is organized as follows: *Section 2* describes the literature review and hypothesis development. *Section 3* presents the methodology of Benford's Law. *Section 4* describes the sample selection. *Section 5* presents the Benford's Law analysis. *Section 6* concludes the study.

2. Literature Review and Hypothesis Development

The manipulations of accounting items (frauds) have been a serious problem for the global financial system. The research related to development of new fraud detection tools has been prevalent in last three decades. Earnings management models like Healy model (1985), Jones model (1991), Modified Jones model (Dechow et al., 1995), performance-based model (Kothari et al., 2005), cash flow model (Dechow & Dichev, 2002), and Roychowdhury model (2006) have been used to detect accrual-based earnings as well as real earnings management. Similarly, for quick detection of frauds, some hand tools like Beneish M-score (Beneish, 1999) and Millers ratio (Miller, 2011) have been developed.

More recently, Benford's Law is used to detect the accounting frauds. This law was developed by Frank Benford in 1938. Since 1938, this law has been applied in various fields like optimizing computer algorithms (Gent & Walsh, 2001) to testing eBay auctions (Gilles, 2007), to detect accounting manipulation (Nigrini, 1996). Enormous numbers of papers have been published in various fields using Benford's law, and the validity of Benford's Law has been proved. Carslaw (1988) performed the second-digit analysis of the reported earnings of New Zealand companies and found more zeros and fewer nines than expected frequencies. The data did not confirm to the expected distribution (Benford's law). Firms tend to round up reported earnings in case those are just below the psychological boundaries.

Using data from Tax returns on the USA Internal Revenue Service's Individual Tax Models Files, Nigrin (1996) found that the low-income taxpayers tend to use more unplanned evasion than high-income taxpayers. Nigrini and Miller (2009) improved the use of Benford's Law performing second order test. In their tests, they used the four sets of transactional data like corporate accounting payables, corporate journal entries, franchisor restaurant's monthly sales and franchisor restaurant revenues and costs. The first three data sets like corporate accounts payable data, corporate journal entry and franchisor monthly sales conform to Benford's law. However, the Franchisor restaurant revenues and cost did not conform to Benford's law. They suggest that the Benford's second-digit test can be used for detection of fraud. Durtschi et al. (2004) explored the data that are likely to conform to Benford's Law and the conditions under which a Benford's Law is likely useful. They found that when the sets of number come from a mathematical combination of numbers (e.g. accounts receivable, accounts payable), transactional level data (e.g. sales, expenses), on large data sets, and the mean of the data set is greater than median and positive skewness, there is an indication of non-conformity of Benford's law. Grabinski and Paszek (2013) applied Benford's Law to large financial data sets of European public listed companies. They used accounting items and financial ratios like net profit, total assets, sales, equity, ROA, ROE, and ROS for investigation purpose. Two stage analyses like first digit and second digit analysis were used. Based on the analysis, accounting items reported by public listed companies are found to be consistent with Benford's Law and financial ratios are consistent to a lesser extent. Asllani and Naco (2014) analyzed the local Albanian hospital data using Benford's law. Using first digit analysis, they found that most of the hospital items like monthly phone charges, rent expenses, etc. deviate from expected distribution. Alali and Romero (2013) provided the evidence of large sample (47431 firm-year observations) of USA public companies, breaking data into multiple periods like Pre-SOX (Note 5) period, SOX-1(2002-2003),

SOX-2(2004-2006), SOX-3(2007), Crisis 1 period (2008) and Crisis 2 period (2009-2010). In their study, they used the financial accounting data like total assets, current assets, inventory, receivables, thecost of goods sold, current liabilities and property, plant and equipment and performed first-digit and first-two digit analysis. They suggested that first digit is an initial test of reasonableness while the first two digit analysis is a more effective test of deviation from Benford's Law to identify potential manipulation. They find that Current liabilities, property, plant and equipment, thecost of goods sold, inventory, receivables and current assets have significant deviation from Benford's Law. The overall study shows that likely overstatement and understatement exist during the analysis period.

Based on the discussions, the present study uses select financial statement items [total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, the cost of goods sold, cash, EBIT, direct tax and indirect tax] of Indian firms'based on their published Profit and Loss accounts and Balance sheets for examining the reliability of reporting numbers using Benford's Law distribution. The following hypothesis is proposed for the purpose.

H_1: The pattern of the numbers of total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, cost of goods sold, cash, EBIT, direct tax and indirect tax do not conform Benford's law.

3. Methodology

The study uses the financial statement items like total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, cost of goods sold, cash, EBIT, direct tax and indirect tax for examining the reliability using Benford's Law distribution. The expected frequencies in the digit of a list of numbers are known as Benford's Law. The numbers starting low first digits (such as 1, 2 and 3) occur more frequently than the numbers starting with high first digits (such as 7, 8 and 9) (Nigrini, 2009). The first digit of the number is the left most digit, and 0 is inadmissible for the first digits. For example in the cases of digits like 28054, 0.0034, the first digits are 2 and 3 respectively. Similarly, second digits are 8 and 4. The first two digits are 28 and 34. According to the Benford's Law, the calculation of probability distribution of any non-zero first digit, second digit, and first two digits are shown as follows.

For first digit of the number

$$P(D_1) = \text{Log}(1 + \frac{1}{D_1})$$

Where: P is the probability of the number D_1, and

D_1 is any number in the set {I, 2, 3,...,9}

For second digit of the number

$$P(D_2) = \sum_{D_{1=1}}^{9} \text{Log}(1 + \frac{1}{D_2})$$

Where: p is the probability of number D_2 and

D_2 is any number in the set $\{0, 1, 2, \ldots.,9\}$

For first two digit of the number

$$P\,(D_1D_2) = d_1d_2) = Log\,(1+\frac{1}{d_1d_2})$$

Where: p is the probability of number D_1D_2 and

d_1d_2 is any number in the set $\{10, 11, 12, \ldots.,99\}$. Table 2 shows the expected frequencies for all digits 0 through 9 in each of the first four places in any number.

Table 1. Expected frequencies based on Benford's law

Digit	Ist place	2nd place	3rd place	4th place
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.19882	0.10097	0.1001
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.0994	0.09994
7	0.05799	0.09035	0.09902	0.0999
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.0850	0.09827	0.09982

Note: The number 143 has three digits, with a 1 as the first digit, 4 as the second digit, and a 3 as the third digit. The table indicates that under Benford's Law the expected proportion of numbers with first digit 1 is 0.03103, the expected proportion of number with a second digit 4 is 0.10031, and the expected proportion of number with third digit 3 is 0.10057.

Source: Nigrini, 1996.

Our study uses test statistics such as Z-statistic and Chi-square statistic for examining the statistcial significnat. For individual distribution, Z-statistics is used and for whole distribution Chi-square test statistics is used.

The difference can be there between observed frequency and expected frequency under Benford's law. Nigrini (1996), Durtschi et al. (2004), and Fleiss (1981) suggest Z-statistic to test the deviations of expected frequency from observed frequencies of individual distribution. The formula of Z-statistic is as follow:

$$Z = \frac{(|P - P_0|) - \frac{1}{2n}}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$

Where p is the observed frequency

P_0 is the expected frequency under Benford's Law; n is the sample size (the term $\frac{1}{2n}$ is a correction term and should be usedonly when it is smaller than ($|P - P_0|$)

If the values of Z-statistic exceed the critical value 1.96, and 2.57, the null hypothesis is rejected at 5 and 1 percent level, respectively.

To test the statistical significance of the whole distribution of observed frequency of first digit, second digit and first two digit against its expected frequency under Benford's Law, Chi-square test statistic with 8, 9 and 89 degrees of freedom are used.

For whole distribution of first digit with 8 degree of freedom. The formula of Chi-sqaure test statistics as follow:

$$\chi^2 = \sum_{D_{n=1}}^{9} \frac{(P - P_0)^2}{P_0}$$

Where: D_n is any number in the set $\{1, 2, 3,\dots,9\}$;

P is the observed frequency

P_0 is the expected frequency

n is the sample size.

For whole distribution of second digit with 9 degree of freedom. The formula of Chi-sqaure test statistics as follow:

$$\chi^2 = \sum_{D_{n=0}}^{9} \frac{(P - P_0)^2}{P_0}$$

Where: D_n is any number in the set $\{0, 1, 2, 3,\dots,9\}$;

P is the observed frequency

P_0 is the expected frequency

n is the sample size.

For whole distribution of first two digit with 89 degree of freedom. The formula of Chi-sqaure test statistics as follow:

$$\chi^2 = \sum_{D_{n=10}}^{90} \frac{(P - P_0)^2}{P_0}$$

http://jsss.macrothink.org

Where: D_n is any number in the set {10, 11, 12,…..,90};

P is the observed frequency

P_0 is the expected frequency

n is the sample size.

If the values of Chi-square test statistic exceed the critical values 15.51 and 20.09 with 8 degrees of freedom, the data is significant at 5 percent and 1 percent level respectively. Similarly, if the Chi-square test statistics exceed the critical values 19.92 and 21.67 with 9 degrees of freedom, the data is significant at 5 percent and 1 percent level respectively. The data is significant at 5 percent and 1 percent level when the Chi-square test statistics exceed the critical values 112.02 and 135.98 with 89 degrees of freedom.

4. Data Collection

The financial statements data are obtained from Center for Monitoring Indian Economy (CMIE), Prowess database in India. The data are collected for years from 2000 to 2014. The initial sample 34,346 firm-year observations are considered for total sample analysis. Further total sample is categorized into two parts (a) business group firms and (b) standalone firms. The sample included 23340 firm-year observations for business group firms and 13,770 firm-year observations for standalone firms. Any data with a value of zero is not considered. *Table 2* provides the firm-year observations used for analysis under Benford's law. "*A business group is a set of firms which, though legally independent, are bound together by a constellation of formal and informal ties and are accustomed to taking coordinated action*" (Khanna & Rivkin, 2001). Companies not classified as business group firm are classified as Standalone firm.

5. Analysis of Result

In this section we report the result of Benford's Law analysis of test variables. The test variables such as total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, cost of goods sold, cash, EBIT, direct tax and indirect tax. The "first digit analysis," "second digit analysis" and "first two digit analyses" are reported using Z-statistics and chi-square test statistics respectively. For individual digit distrubution, Z test statistics is used and for whole distribution, chi-square test statistics is used. Z-statistics is untabulated (Note 6).

Table 3 reports the descriptive statistics of the total sample show the following: average total assets are 7115.80 million with a median company size 746.70 million; average fixed assets 5601.10 million; average property, plant and equipment invested by company are 4311.70 million; average current assets are 9736.60 million; with average receivables of 992.36 million; with average inventory of 3543.70 million and average cash of 34.69 million; average revenues are 11464.00 million; with average cost of goods sold of 3553.70 million; average current liabilities are 3546.10 million; average earnings before and interest and taxes are 5247.30 million; average selling and distribution expenses are 493.17 million; average direct tax and indirect tax are 284.41 million and 915.80 million respectively. Following Wallace (2002) and Durtschi et al. (2004), we find that mean of test variables are larger than the median and skewness value is positive. So our data set is suitable for Benford's Law analysis.

5.1 First-Digit Analysis

To obtain the initial test of reasonableness, we examine the first digit analysis of test variables such as total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, cost of goods sold, cash, EBIT, direct tax and indirect tax. Table 3 reports that total assets, property, plant and equipments, current liabilities, sales, cash and direct tax have significant deviation from Benford's law. Rest of variables such as receivables, fixed assets, inventory, current assets, selling and distribution expenses, cost of goods sold, EBIT and indirect tax confirm the Benford's Law.

Table 2. The description of data of balance sheet (B/S) and Profit & Loss (P&L) items

	Total sample	Business group firm	Standalone firm
No of Companies	2474	1556	918
No of firm-year observations	34636	23340	13770
Data deleted			
(a) equal to zero	867	540	327
(b) less than 0 and > -10		0	
Valid data used in this research			
(a) Total assets	17839	14067	10045
(b) Receivables	15943	12275	8694
(c) Fixed assets	19570	15051	8789
(d) PPE	19570	15051	8789
(e) Inventory	2690	3434	1947
(f) Current assets	2842	3597	2072
(g) Current liabilities	19528	15030	8752
(h) Sales	18617	14379	8284
(i) Selling and distribution expenses	17455	13499	7733
(j) Cost of goods sold	19244	15051	8789
(k) Cash	16165	12516	7182
(l) EBIT	18893	14633	8480
(m) Direct Tax	13518	10585	6154
(n) Indirect Tax	17555	13605	7790

5.2 Second-Digit and First two-Digit Analysis

Nigrini (2009) find the second order test can detect (a) anomalies occurring in data

downloads, (b) rounded data, (c) transaction level data in more accuracy. Given the existence departures from Benford's Law for the first digit analysis, we perform the second digit analysis and first two digit analyses of test variables.

As per the second digit analysis (Table 3), total assets, receivables, fixed assets, property, plant and equipment, current liabilities, sales, selling and distribution expenses, cash, EBIT, direct tax and indirect tax exceed the critical value of chi-square distribution and do not confirm the Benford's Law. The rest of the variables such as inventory, current assets, and cost of goods sold have no significant deviation from Benford's Law distribution. It is also found that most of the variables (Table 3) such as total assets, receivables, fixed assets, property, plant and equipment, current liabilities, selling and distribution expenses, cash, direct tax and indirect which are significant in second digit analysis are also significant in first two digit analysis. However, test variables such as current assets and sales do not exceed the critical value of chi-square distribution in first two digit analysis. The data have no anomalies under Benford's Law distribution.

Table 3. Descriptive statistics of accounting variables

Variables	Mean	Median	Skewness	Std. Dev.	Kurtosis
(a) Total assets	7115.80	746.70	28.98	50799	1197.20
(b) Receivables	992.36	141.10	24.33	5890	822.90
(C) Fixed assets	5601.10	689.25	33.42	43138	1465.10
(d) PPE	4311.70	466.60	28.25	32042	1037.30
(e) Inventory	3543.70	418.00	18.07	22430	405.83
(F) Current assets	9739.60	1142.10	17.34	56945	367.65
(g) Current liabilities	3546.10	312.25	27.30	26536	1060.20
(h) Sales	11464.00	1224.70	27.98	106400	960.83
(i) SG&A expenses	493.17	37.20	25.20	3209	930.74
(j) COGS	3553.70	476.59	28.38	33849	1000.00
(k)Cash	34.69	0.90	16.44	273	339.45
(l)EBIT	5247.30	661.80	28.39	47115	1023.80
(m)Direct Tax	284.41	13.70	16.92	1840	374.43
(n)Indirect Tax	915.80	60.10	21.06	7385	562.36

5.3 Additional Analysis

In this section, following the prior studies, we also examine the deviations from Benford's Law for business group firms and standalone firms (Kim & Yi, 2006; Beuselinck & Deloof, 2014) and for large size firms and small size firms (Ajit et al., 2013; Alali & Romero, 2013). The composite result of data deviation from Benford's Law distribution of additional analysis

are shown in Table 7.

Table 4. Chi-square statistics of first digit, second digit and first two digit distribution of total sample under Benford's law

Variables	First-Digit	Second-Digit	First Two-Digit
(a) Total assets	39.70**	362.09**	820.35**
(b) Receivables	32.76**	111.27**	258.52**
(C) Fixed assets	10.57	23.02**	154.60**
(d) PPE	32.76**	111.27**	258.52**
(e) Inventory	25.36**	10.80	118.50*
(F) Current assets	12.71	14.88	104.25
(g) Current liabilities	19.37*	31.10**	133.10**
(h) Sales	12.59	26.64**	93.96
(i) SG&A expenses	13.62	1054.73**	1254.16**
(j) COGS	24.12**	13.52	154.64**
(k)Cash	454.95**	39170.39**	39494.09**
(l)EBIT	11.94	39.71**	160.16**
(m)Direct Tax	27.78**	4308.40**	4597.49**
(n)Indirect Tax	8.07	1926.17**	2068.70**

(i) The score of chi-square exceeding 15.51, 16.92 and 112.02 are 95% level of significance difference and 8, 9 and 89 degree of freedom respectively and also identified by *.

(ii) The score of chi-square exceeding 20.09, 21.67 and 122.94 are 99% level of significance difference and 8, 9 and 89 degree of freedom respectively and also identified by **

5.3.1 Business Group Firms Versus Standalone Firms

Prior studies suggest that business group firms do more earnings management than standalone firms (Kim & Yi, 2006; Beuselinck & Deloof, 2014). We examine the data anomalies of test variables such as total assets, receivables, fixed assets, property, plant and equipment, inventory, current assets, current liabilities, sales, selling and distribution expenses, cost of goods sold, cash, EBIT, direct tax and indirect tax in business group firms and standalone firms using Benford's Law analysis. Table 5 reports the "first digit analysis," "second digit analysis," and "first two digit analysis" of test variables. In first digit analysis of business group firms, total assets, property, plant and equipment, current liabilities, sales, selling and distribution expenses, cash, EBIT and direct tax exceed the critical value of chi-square test statistics. Given the existence departures from Benford's Law for the first digit analysis, all test variables of chi-square test statistics excluding inventory, sales and current liabilities have significant deviation from Benford's Law distribution in second digit analysis

and first two digit analysis respectively. In standalone firms, total assets, property, plant and equipment, inventory, sales and cash exceed the critical value of chi-square distribution in first digit analysis. Consistent with first digit analysis, chi-square test statistics of total assets, receivables, fixed assets, selling and distribution expenses, cash, direct tax, and indirect tax have significant deviation from Benford's Law distribution.

Table 5. Chi-square statistics of first digit, second digit and first two digit distribution of business group firms and standalone firms under Benford's law

	Business group firm			Standalone firm		
Variables	First digit	Second digit	First two digit	First two digit	Second digit	First two digit
(a) Total assets	29.21**	144.68**	411.84**	18.73*	273.81**	591.22**
(b) Receivables	8.68	109.17**	223.27**	5.66	102.60**	229.58**
(C) Fixed assets	15.12	20.29**	151.13**	5.71	29.73**	183.82**
(d) PPE	16.75*	55.29**	196.44**	35.20**	67.41**	199.52**
(e) Inventory	21.90**	10.49	102.59	16.76*	6.13	125.61**
(F) Current assets	9.87	17.42*	117.29*	14.52	18.38*	103.69
(g) Current liabilities	22.40**	24.25**	104.02	13.81	18.52*	104.00
(h) Sales	20.44**	13.11	103.22	5.37	19.93*	73.84
(i) SG&A expenses	21.26**	360.76**	458.08**	5.89	701.23**	909.27**
(j)COGS	14.03	18.60*	115.67*	4.24	8.35	94.48
(k)Cash	243.59**	14739.50**	14910.77**	241.49**	24440.32**	24698.65**
(l)EBIT	19.81*	18.92*	122.99**	3.85	20.20*	83.01
(m)Direct Tax	21.86**	1630.46**	1876.45**	10.86	2687.13**	2823.83**
(n)Indirect Tax	8.12	770.42**	881.96**	7.08	1163.52**	1285.82**

(i) The score of chi-square exceeding 15.51, 16.92 and 112.02 are 95% level of significance difference and 8, 9 and 89 degree of freedom respectively and also identified by *.

(ii) The score of chi-square exceeding 20.09, 21.67 and 122.94 are 99% level of significance difference and 8, 9 and 89 degree of freedom respectively and also identified by **.

After analyzing the results (Table 7) of the "first digit," "second digit" and "first two digit," it can be concluded that data anomalies are more prominent in the case of business group firms than standalone firms. This result is in support with Carslaw (1988) and Thomas (1989) performed for unusual patterns in earnings number on a sample of 2315 firms taken from the CMIE Prowess database over a six year period (2006-2011) suggests that business affiliated firms are indulged in more data anomalies than non-family counterparts.

5.3.2 Large Size Versus Small Size Firms

Securities Exchange Board of India (SEBI) Development Research Group (DRG) study by

Ajit, Malik and Verma (2013) finds that average earnings management by India corporate sector is 2.9% of the total asset which is comparable to companies from US, Europe and elsewhere in the world. In their study, they find that small size firms indulge in earnings management more than large size firms. Using Benford's Law analysis, (Johnson, 2009) find that smaller firms have more data anomalies than larger firms. However, Alali and Romero (2013) find that large firms have significant deviation from Benford's Law in data anomalies. Following the above studies, we classify the total assets values, and identifying the top-50 percent observations as large firms and the bottom-50 percent as small as firm.

Table 5 reports the "first digit", "second digit" and "first two digit" analysis of test variables of large size firms and small size firms. In first digit analysis of large size firms, total assets, receivables, and cash exceed the critical value of chi-square test statistics. Consistent with first digit analysis, chi-square test statistics of variables such as total assets, fixed assets, property, plant and equipment, selling and distribution expenses, cash, EBIT, direct tax and indirect tax have significant deviation from Benford's Law distribution in second digit analysis and first two digit analysis respectively. In first digit analysis of small size firms, test variables such as total assets, receivables, Property, plant and equipment, inventory, cash and direct tax exceed the critical value of chi-square test statistics. In second digit analysis total assets, receivables, Property, plant and equipment, current liabilities, sales, cash, EBIT, direct tax and indirect tax have significant deviation from Benford's Law distribution. While in the first two digit analysis, test statistics of total assets, receivables, fixed assets, inventory, selling and distribution expenses, cash, direct tax and indirect tax exceed the critical values of chi-square test. Comparing the "first digit," "second digit" and "first two digit analyses," we find that smaller firms have more data anomalies than larger firms. The comparative analysis is shown in Table 7.

Table 6. Chi-square statistics of first digit, second digit and first two digit distribution of small size firm and large size firm under Benford's law

	Small size firms			Large size firms		
	First digit	Second digit	First two digit	First digit	Second digit	First two digit
Variables						
(a) Total assets	791.10**	586.15**	2320.63**	572.16**	30.66**	781.64**
(b) Receivables	35.97**	347.36**	526.24**	32.77**	12.01	134.32**
(C) Fixed assets	4.37	6.24	103.53	11.16	24.99**	126.97**
(d) PPE	16.85*	28.49**	118.50*	10.60	62.83**	143.58**
(e) Inventory	16.19*	7.90	119.94*	11.52	4.30	70.63
(F) Current assets	9.80	10.76	110.38	6.59	6.64	93.28
(g) Current liabilities	7.16	18.42*	84.41	6.51	12.54	100.13
(h) Sales	8.86	18.06*	93.61	6.62	11.44	83.92
(i) SG&A expenses	5.17	298.05**	400.52**	7.52	319.72**	443.47**

(j)COGS	9.13	11.51	82.37	8.39	5.69	99.83
(k)Cash	138.79**	12325.52**	12456.29**	166.78**	11436.04**	11618.76**
(l)EBIT	10.10	25.76**	86.16	6.07	23.46**	114.23*
(m)Direct Tax	22.19**	1235.77**	1436.49**	8.75	1420.38**	1570.28**
(n)Indirect Tax	15.01	625.47**	705.99**	2.26	598.75**	716.01**

(i) The score of chi-square exceeding 15.51, 16.92 and 112.02 are 95% level of significance difference and 8, 9 and 89 degree of freedom respectively and also identified by *.

(ii) The score of chi-square exceeding 20.09, 21.67 and 122.94 are 99% level of significance difference and 8, 9 and 89 degree of freedom respectively and also identified by **.

Table 7. Composite results of digital analysis

variables	Business group firm			Standalone firm			Small size firm			Large size firm		
Digit	First	Second	First two	First	Second	First two	First	Second	First two	First	Second	First two
(a) Total assets	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(b) Receivables		✓	✓		✓	✓	✓	✓	✓	✓		✓
(C) Fixed assets		✓	✓		✓	✓					✓	✓
(d) PPE	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
(e) Inventory	✓			✓		✓	✓		✓			
(F) Current assets		✓	✓		✓							
(g) Current liabilities	✓	✓			✓			✓				
(h) Sales	✓			✓				✓				
(i) SG&A expenses	✓	✓	✓		✓	✓		✓	✓		✓	✓
(j)COGS				✓	✓							
(k)Cash	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(l)EBIT	✓	✓	✓		✓	✓		✓			✓	✓
(m)Direct Tax	✓	✓	✓		✓	✓	✓	✓	✓		✓	✓
(n)Indirect Tax		✓	✓		✓	✓		✓	✓		✓	✓

Note: The √ indicates the variables do not confirm the Benford's Law distribution.

6. Conclusion

Irrespective of the distribution of the underlying data, Benford's' law test is useful in the hands of auditors to find out the data anomalies before auditing. This law helps the internal auditing with providing assurance that fraud prevention and detections controls are adequate and are functioning as designed. The deviation between observed frequency and expected frequency means signal of data anomalies according Benford's Law distribution. Using large sample of Indian public listed companies, we conduct an analysis of the "first digit," "second digit" and "first two digit analyses" of test variables that are used in research for measuring the discretionary accruals, abnormal expenditures. In total sample we find that most of the variables have significant deviation from Benford's Law distribution. Further analysis shows that business group firms indulge in more data anomalies than standalone firms. Small size firms have more data anomalies than large size firms.

It is pertinent to note here that, as an emerging economy, India's capital market is well developed with a large number of listed companies and attracts significant foreign portfolio investment. Irrespective of these attributes, there may be some loopholes that companies take advantage and manipulate the numbers. This study recommends that, in addition to independent audit committee, there should be an additional institutional mechanism for verification of financial reports. It is high time, rotation of auditors every two/three years is implemented. Stakeholders, particularly lender/bankers need to more vigilant with thepresence of nominee directors.

Benford's Law does not detect the relationship between variables. Rather it detects the irregular distribution of numbers. Though it is used as a hand tool for detecting a potential problem before exploring the detailed information, other statistical methods must be used to corroborate the results found.

Acknowledgements

References

Ajit, D., Malik, S., & Verma, V. K. (2013). SEBI DRG Study Earnings Management in India. working paper, DRG SEBI, India. Retrieved from http://www.sebi.gov.in/cms/sebi_data/DRG_Study/EMiM.pdf

Alali, F. A., & Romero, S. (2013). Benford's Law: Analyzing a decade of financial data. *Journal of Emerging Technologies in Accounting, 10*(1), 1-39. http://dx.doi.org/10.2308/jeta-50749

Asllani, A., & Naco, M. (2014). Using Benford's Law for Fraud Detection in Accounting Practices. *Journal of Social Science Studies, 2*(1), 129. http://dx.doi.org/10.5296/jsss.v2i1.6395

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal, 55*(5), 24-36. http://dx.doi.org/10.2469/faj.v55.n5.2296

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American*

Philosophical Society, 551-572.

Beuoselinck, C., & Deloof, M. (2014) Earnings management in business groups: tax incentives or expropriation concealment? *The international Journal of Accounting, 49*(1), 27-52. http://dx.doi.org/10.1016/j.intacc.2014.01.008

Carslaw, C. A. (1988). Anomalies in income numbers: Evidence of goal oriented behavior. *Accounting Review,* 321-327.

Coderre, D. G. (2000). Computer assisted fraud detection. *Internal Auditor, 57*(4), 25-25.

Dechow, P. M., & Dichev, I. D. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The accounting review, 77*(s-1), 35-59. http://dx.doi.org/10.2308/accr.2002.77.s-1.35

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. *Accounting review,* 193-225.

Diekmann, A. (2007). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Journal of Applied Statistics, 34*(3), 321-329. http://dx.doi.org/10.1080/02664760601004940

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of forensic accounting, 5*(1), 17-34.

Fleiss, J. L. (1981), *The Measurement of Interrater Agreement, Statistical Methods for Rates and Proportions* (Second Edition), John Wiley & Sons, Inc., New York 212-304.

Gent, I., & Walsh, T. (2001). Benford's law. APES-25-2001. Research Reports of the APES Group. United Kingdom.

Giles, D. E. (2007). Benford's law and naturally occurring prices in certain eBay auctions. *Applied Economics Letters, 14*(3), 157-161. http://dx.doi.org/10.1080/13504850500425667

Grabiński, K., & Paszek, Z. (2013). Examining reliability of large financial datasets using Benford's law. *Economic Themes, 51*(3).

Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics, 7*(1), 85-107. http://dx.doi.org/10.1016/0165-4101(85)90029-1

Jaiswal, M., & Banerjee, A. (2011) Study on the state of corporate governance in India.

Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research,* 193-228. http://dx.doi.org/10.2307/2491047

Khanna, T., & Rivkin, J. W. (2001). Estimating the performance effects of business groups in emerging markets. *Strategic management journal, 22*(1), 45-74. http://dx.doi.org/10.1002/1097-0266(200101)22:1<45::AID-SMJ147>3.3.CO;2-6

Kim, J-B., & Yi, C. H. (2006). Ownership structure, business group affiliation, listing status, and earnings management: evidence from Korea. *Contemporary Accounting Research, 23*(2),

427-464. http://dx.doi.org/10.1506/7T5B-72FV-MHJV-E697

Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of accounting and economics, 39*(1), 163-197. http://dx.doi.org/10.1016/j.jacceco.2004.11.002

Miller, J. E., & Miller, S. E. (2011). The Miller Ratio: Is It Really This Simple? *Journal of Business & Economics Research (Online), 9*(12), 45. http://dx.doi.org/10.19030/jber.v9i12.6605

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics, 4*(1), 39-40. http://dx.doi.org/10.2307/2369148

Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law. *The Journal of the American Taxation Association, 18*(1), 72.

Nigrini, M. J., & Miller, S. J. (2009). Data diagnostics using second-order tests of Benford's law. *Auditing: A Journal of Practice & Theory, 28*(2), 305-324. http://dx.doi.org/10.2308/aud.2009.28.2.305

Raimi, R. A. (1976). The first digit problem. *The American Mathematical Monthly, 83*(7), 521-538. http://dx.doi.org/10.2307/2319349

Rauch, B., Göttsche, M., Brähler, G, & Engel, S. (2011). Fact and Fiction in EU‐Governmental Economic Data. *German Economic Review, 12*(3), 243-255. http://dx.doi.org/10.1111/j.1468-0475.2011.00542.x

Notes

Note 1. http://www.businesstoday.in/magazine/features/vishal-retail-transformation-to-v2-benefits/story/21755.html

Note 2. http://articles.economictimes.indiatimes.com/2011-08-25/news/29926869_1_r-subramanian-subhiksha-renuka-ramnath

Note 3. http://www.business-standard.com/article/companies/reebok-india-stands-by-its-rs-870-cr-fraud-allegations-112121400126_1.html

Note 4. http://articles.economictimes.indiatimes.com/2012-03-01/news/31113571_1_dlf-veritas-negative-research-report

Note 5. The Sarbanes-Oxley Act of 2002 (often shortened to SOX) is legislation passed by the U.S. Congress to protect shareholders and the general public from accounting errors and fraudulent practices in the enterprise, as well as improve the accuracy of corporate disclosures

Note 6. The results will be furnished on request.

Copyright Disclaimer